Figure 4 shows a multiple alignment of *E. Coli* DXPR (SEQ ID NO:1) to *S. aureas* homoserine dehydrogenase (1EBF_A; SEQ ID NO:2).

Please replace the second paragraph on page 11 (lines 14-29) with the following:

As used herein, the term "ligand" refers to a molecule that can specifically bind to a polypeptide. Specific binding, as it is used herein, refers to binding that is detectable over non-specific interactions by quantifiable assays well known in the art such as those that measure association rates, dissociation rates or equilibrium association or dissociation constants. A ligand can be essentially any type of natural or synthetic molecule including, for example, a polypeptide, nucleic acid, carbohydrate, lipid, amino acid, nucleotide or any organic derived compound. The term also encompasses a cofactor or a substrate of a polypeptide having enzymatic activity, or substrate that is inert to catalytic conversion by the bound polypeptide. Specific binding to a polypeptide can be due to covalent or non-covalent interactions.

Please replace the paragraph spanning pages 17 and 18 (page 17, line 22, through page 18, line 11) with the following:

The methods can also be used with a set of amino acid sequences that are preselected for a particular structural or functional characteristic. A preselected range of structural or functional characteristics for a set of polypeptides used in the methods can include, for example, binding to a particular ligand, interacting with a particular biological component such as another protein, common enzymatic function, common structural motifs or folds, common subcellular localization or co-expression due to a particular stimulus or developmental or growth stage. Those skilled in the art will be able to preselect a set of amino acid sequences based on that which is known for particular sequences as provided in the scientific literature or in annotations of particular databases. Examples of subsets of polypeptides from which subsets can be identified in the methods of the invention include, for example, kinases, G-protein coupled receptors, nuclear factors, proteases, dehydrogenases, phosphatases, transcription factors, nucleotide binding enzymes or membrane proteins.

Please replace the paragraph spanning pages 19 and 20 (page 19, line 15, through page 20, line 8) with the following:

A set of amino acid sequences used in the methods can be translated from one or more nucleic acid sequences in a nucleic acid sequence database. Accordingly, the methods can include a step of translating the coding regions of a nucleic acid sequence into amino acid sequences. A coding region of a nucleic acid sequences can be translated according to the appropriate genetic code for the organism from which the nucleic acid sequence is derived. The coding region can be a predetermined portion of the sequence or in the case where exons and introns are present a predetermined set of spliced portions identified, for example, from annotations of the nucleic acid in the database. Alternatively, the coding region can be predicted or determined based on methods known in the art for predicting gene structure or coding sequence location. Computational methods for predicting the coding region of a nucleic acid sequence are known in the art as described in Pevzner, Computational Molecular Biology, an Algorithmic Approach, The MIT Press, Cambridge MA (2000), and include, for example, statistical approaches based on codon usage or in-frame hexamer count, similarity based approaches, spliced alignment approaches and Hidden Markov based approaches such as GENSCAN.

Please replace the paragraph spanning pages 21 and 22 (page 21, line 16, through page 22, line 3) with the following:

The dynamic programing algorithm is a mathematically rigorous method of pairwise sequence comparison and can be used according to several variants including, for example, Needleman-Wunsch (Needleman and Wunsch, J. Mol. Biol. 48:443-453 (1970)), Sellers (Sellers, J. Appl. Math. 26:787-793 (1974)), quasi-global alignment (Sellers Proc. Natl. Acad. Sci. USA 76:3041-3041 (1979)) and Smith-Waterman (Smith and Waterman, J. Mol. Biol. 147:195-197 (1981) and Waterman and Eggert, J. Mol. Biol. 197:723-728 (1987)). The dynamic programming algorithm is rigorous and therefore, well suited for finding optimum alignments and sequence comparison scores for a set of amino acid sequences. The dynamic programing algorithm, being rigorous is also computationally demanding. In applications of the methods in which large sequence sets are used or less rigorous comparison is required a heuristic search algorithm can be used.

Please replace the second paragraph on page 22 (lines 4-28) with the following:

Heuristic algorithms that can be used in the methods of the invention include, for example, BLAST and FASTA. BLAST, Basic Local Alignment Search Tool, uses a heuristic algorithm that reduces the computational requirements of the Smith-Waterman algorithm by seeking local alignments prior to comparing sequences in a restricted version of the Smith-Waterman algorithm. BLAST is therefore able to detect relationships among sequences including those which share only isolated regions of similarity including, for example, protein domains (Altschul et al., J. Mol. Biol. 215:403-410 (1990)). BLAST divides sequences into a list of overlapping words and extends the list to include all words that score above a predefined matrix-defined threshold. This threshold limits the number of matches that will be passed from the heuristic screening step to the comparison step. Those skilled in the art can use BLAST according to a default parameters as described by Tatiana et al., FEMS Microbial Lett. 174:247-250 (1999) or on the National Center for Biotechnology Information web page at ncbi.nlm.nih.gov/BLAST/. Alternatively, parameters such as the length of the words, value of the predefined matrix-defined threshold or type of similarity matrix utilized can be adjusted to suit a particular application of the methods of the invention.

Please replace the paragraph spanning pages 23 and 24 (page 23, line 17, through page 24, line 2) with the following:

FASTA uses a word search algorithm as a heuristic screen prior to performing a restricted Smith-Waterman alignment (Pearson and Lippman, Proc. Natl. Acad. Sci. USA 85:2444-2448 (1988)). In the word search both the query and library sequences are divided into overalapping words of specified length. The lists of words for the query and library sequences are compared in a matrix and the diagonal with the most matching words is taken as the region most likely to contain the best alignment. The results from the word search are used to identify sequences with sufficient similarity to use in the subsequent alignment step. Those skilled in the art can use default parameters or adjust parameters such as word size, window size for defining the length of insertions or deletions one sequence can accumulate relative to another or the type of similarity matrix utilized.

Please replace the paragraph spanning pages 26 and 27 (page 26, line 26, through page 27, line 14) with the following:

The distance between two similarity signatures, that are represented as a first and second vector in high dimensional space, can be determined based on the distances separating the points of the first vector from the points of the second vector. A variety of distance measures are known in the art can be used in the methods of the invention including, for example, Euclidian distance. Euclidian distance is the square root of the sum of the difference between each of the elements in the two compared vectors, squared. Another distance is the Mahalanobis distance, which scales the difference in each coordinate by the inverse of the variance in that dimension as described, for example, in

Mahalanobis, Proc. Natl. Acad. Sci. USA 12:49-55 (1936). The cosine of the angle between the two vectors can also be computed and used as a distance metric. Hamming distance between two vectors is also useful in the methods of the invention and it is given by the count of the number of elements in which the two vectors differ.

Please replace the paragraph spanning pages 27 and 28 (page 27, line 15, through page 28, line 2) with the following:

Distances that are particularly useful when binary sequence comparison scores are used include, for example, the exclusive OR which is a reduction of a hamming distance to a binary case, again being a count of the number of elements differing between the two vectors that are compared. The Tanimoto coefficient is the ratio of bits set (where a bit set is a bit that is equal to 1) for both vectors to the total number of bits set in either vector. A generalization of the Tanimoto coefficient is the Tversky Similarity, where both vectors can be given different weighting as described in Sneath and Sokal, Numerical Taxonomy WH Freeman, San Francisco (1973). Those skilled in the art will recognize that this is only a partial list of the methods known in the art for measuring distance between vectors and will be able to use other known methods for measuring distance between vectors in the methods to determine the distance between sequence similarity signatures according to the teaching herein.

Please replace the paragraph spanning pages 33 through 35 (page 33, line 27, through page 35, line 7) with the following:

Common structural properties can be identified by comparing the three dimensional structures of two or more polypeptides or a bound ligand using methods known in the art including, for example, cluster analysis of structures, visual inspection and pairwise structural comparisons. Cluster analysis of structures is commonly performed by, but not limited to, partitioning methods or hierarchical methods as described, for example, in Kauffman and Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley and Sons Inc., New York (1990). Partitioning methods that can be used include, for example, partitioning around medoids, clustering large applications, and fuzzy analysis, as described in Kauffman and Rousseeuw, *supra*. Hierarchical methods useful in the invention include, for example, agglomerative nesting, divisive analysis, and monothetic analysis, as described in Kauffman and Rousseeuw, *supra*. Algorithms for cluster analysis of molecular structures are known in the art and include, for example, COMPARE (Chiron Corp, 1995; distributed by Quantum Chemistry program Exchange, Indianapolis IN). COMPARE can be used to make all possible pairwise comparisons between a set of conformations of polypeptides or bound ligands or portions thereof. COMPARE reads PDB files and uses a Ferro-Hermanns ORIENT algorithm for a least squares root mean square (RMS) fit. The structures can be clustered into groups using the Jarvis-Patrick nearest neighbors algorithm. Based on the RMS deviation between polypeptide structures or bound conformations of a ligand, or portions thereof, a list of 'nearest neighbors' for each structure is generated. Two structures are then grouped together or clustered if: (1) the RMS deviation is sufficiently small and (2) if both structures share a determined number of common 'neighbors'. Both criteria are adjusted by the program to generate clusters based on a user defined cutoff for distance between individual clusters. Follow up analysis can be conducted using InsightII to verify structural clusters. Thus, two or more polypeptides can be confirmed as being in the same cluster or a polypeptide can be assigned to one of two or more proximal clusters based on common cluster assignment evaluated by both sequence based clustering and structure-based clustering.

Please replace the paragraph spanning pages 35 and 36 (page 35, line 25, through page 36, line 22) with the following:

Using methods such as those described above, one skilled in the art will know how to identify structures that are substantially the same. For example, similarity can be evaluated according to the goodness of fit between two or more three-dimensional models of a polypeptide or bound ligand, or fragments thereof. Goodness of fit can be represented by a variety of parameters known in the art including, for example, the root mean square deviation (RMSD). A lower RMSD between structures correlates with a better fit compared to a higher RMSD between structures (see for example, Doucet and Weber, Computer-Aided Molecular Design: Theory and Applications, Academic Press, San Diego, CA (1996)). Polypeptides having substantially the same structures can be identified by comparing mean RMSD values for the backbones of the polypeptides. Polypeptides, or fragments thereof, having substantially the same structures can have a mean backbone RMSD compared to each other that is less than about 5 Å or less than about 3 Å. Those skilled in the art will know that despite a high RMSD between overall structures indicating overall structural differences, two polypeptides can contain domains or other regions that are similar. Thus, a model used in comparing polypeptide structures can be that of the backbone structure of a domain or other region of the polypeptide. Bound conformations of a ligand having substantially the same structures can have a mean RMSD compared to each other that is less than about 1.1 Å.

Please replace the second paragraph on page 39 (lines 7-13) with the following:

The invention can be used with any ligand that binds to two or more different polypeptides having different sequences including, for example, chemical or biological molecules such as simple or complex organic molecules, metal-containing compounds, carbohydrates, peptides, peptidomimetics, carbohydrates, lipids, nucleic acids, and the like.

Please replace the third paragraph on page 39 (lines 14-26) with the following:

In one embodiment, the methods of the invention can be used with a ligand that is a nucleotide derivative including, for example, a nicotinamide adenine dinucleotide-related molecule.

Nicotinamide adenine dinucleotide-related (NAD-related) molecules that can be used in the methods of the invention can be selected from the group consisting of oxidized nicotinamide adenine dinucleotide (NAD$^+$), reduced nicotinamide adenine dinucleotide (NADH), oxidized nicotinamide adenine dinucleotide phosphate (NADP$^+$), and reduced nicotinamide adenine dinucleotide phosphate (NADPH). An NAD-related molecule can also be a mimetic of the above-described molecules.

Please replace the paragraph spanning pages 46 and 47 (page 46, line 9, through page 47 line 2) with the following:

In yet another representation, the conformer model can be a volume surrounding all or a subset of the bound conformations of a ligand bound to polypeptides in a cluster. A model showing volume can be useful for comparing other structures in a fitting format such that a structure which fits within the volume of the model can be identified as substantially similar to the model. One approach that can be used to fit a structure to a volume is comparison of equivalent surface patches using gnomonic projection as described for example in Chau and Dean, J. Mol. Graphics 7:130 (1989). Use of a gnomonic projection to compare structures is also described in Doucet and Weber, Computer-Aided Molecular Design: Theory and Applications, Academic Press, San Diego CA (1996). Algorithms which can be used to fit a structure to a volume are known in the art and include, for example, CATALYST (Molecular Simulations Inc., San Diego, CA) and THREEDOM which is a part of the INTERCHEM package which makes use of an Icosahedral Matching Algorithm (Bladon, J. Mol. Graphics 7:130

(1989)) for the comparison and alignment of structures. Methods of identifying a binding compound by searching a database of structures using a gnomonic projection are described, for example, in U.S. patent application number 09/753,020, which is hereby incorporated by reference.

Please replace the paragraph spanning pages 78 and 79 (page 78, line 15, through page 79, line 12) with the following:

Sequence comparison signatures were determined for the NAD(P)-binding sequences (including 28 DXPR sequences) in the Swiss-Prot database and clustering was performed as described in Examples I and II. The 28 DXPR sequences formed one cluster. When visualized in a comparison matrix, the DXPR cluster was proximal to other clusters. These other clusters were composed of aspartate semialdehyde dehydrogenase, homoserine dehydrogenase, N-acetyl-g-glutamyl phosphate reductoisomerase, or glyceraldehyde 3-phosphate dehydrogenase; all of which share a common NAD(P)-binding Rossmann fold. The proximity correlated with local sequence identity between DXPR sequences and sequences of these other clusters, ranging from about 17 to 40% local sequence identity. Although the E-scores of these sequence identities were between 0.1 and 2.0, these clusters were identified as related groups because multiple DXPR sequences systematically showed cross-talk to only the above mentioned sequence clusters. In particular, cross-talk was identified as low sequence identity (less than 30%) between the cluster containing DXPR and a few sequences belonging to other clusters, which showed a pattern that was distinct from a pattern observed